DOCUMENT RESUME

ED 068 441 SP 005 893

AUTHOR	Dalkey, Norman C.; Rourke, Daniel L.
TITLE	Experimental Assessment of Delphi Procedures with Group Value Judgments.
INSTITUTION	Rand Corp., Santa Monica, Calif.
SPONS AGENCY	Advanced Research Projects Agency, Washington, D. C.
REPORT NO	R-612-ARPA
PUB DATE	Feb 71
NOTE	58p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Cluster Analysis; *Educational Research; *Group Norms; *Higher Education; *Values

ABSTRACT

This report describes the results of an experiment assessing the appropriateness of Delphi procedures for formulating group value judgments. Two groups of subjects--upperclass and graduate students from UCLA--were asked to generate and rate value categories relating to higher education and the quality of life. The initial lists (300 and 250 items respectively) were aggregated by the experimental teams to 45 and 48 items respectively. The subjects then rated all possible pairs of these items with respect to their similarity. The average similarity ratings were analyzed by a clustering routine. Subjects from both groups made estimates of the relative contributions of each of the educational categories to each of the quality of life categories. The primary data analyses concerned the importance rating. Three aspects were examined: the quality of the distributions of the responses, the correlation between ratings by different groups and different rating techniques, and the amount of change and degree of convergence upon iteration with feedback. The experiment furnished support for the conclusion that Delphi procedures are appropriate for processing value material as well as factual material. The substantive data concerning higher education and individual life also indicated the possibility that current notions of the role of the university are somewhat loosely tied to the basic interests of the students. A 13-item bibliography is included. (Author/MJM)

ED 068441

ARPA ORDER NO. 189-1

R-612-ARPA

February 1971

EXPERIMENTAL ASSESSMENT OF DELPHI PROCEDURES WITH GROUP VALUE JUDGMENTS

Norman C. Dalkey and Daniel L. Rourke

A Report prepared for

ADVANCED RESEARCH PROJECTS AGENCY

FILMED FROM BEST AVAILABLE COPY

Rand
SANTA MONICA, CA. 90406

AD 708 473

This research is supported by the Advanced Research Projects Agency under Contract No. DAHC15 67 C 0141. Views or conclusions contained in this study should not be interpreted as representing the official opinion or policy of Rand or of ARPA.

ARPA ORDER NO.: 189-1

ED 068441

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

R-612-ARPA

February 1971

EXPERIMENTAL ASSESSMENT OF DELPHI PROCEDURES WITH GROUP VALUE JUDGMENTS

Norman C. Dalkey and Daniel L. Rourke

SP 005 853

A Report prepared for

ADVANCED RESEARCH PROJECTS AGENCY

Rand
SANTA MONICA, CA. 90406

THIS DOCUMENT HAS BEEN APPROVED FOR PUBLIC RELEASE AND SALE; ITS DISTRIBUTION IS UNLIMITED

Rand maintains a number of special, subject bibliographies containing abstracts of Rand publications in fields of wide current interest. The following bibliographies are available upon request:

Africa • Arms Control • Civil Defense • Combinatorics
Communication Satellites • Communication Systems • Communist China
Computing Technology • Decisionmaking • East-West Trade
Education • Foreign Aid • Health-related Research • Latin America
Linguistics • Long-range Forecasting • Maintenance
Mathematical Modeling of Physiological Processes • Middle East
Policy Sciences • Pollution • Procurement and R&D Strategy
Program Budgeting • SIMSCRIPT and Its Applications • Southeast Asia
Systems Analysis • Television • Urban Problems • USSR
Water Resources • Weather Forecasting and Control

To obtain copies of these bibliographies, and to receive information on how to obtain copies of individual publications, write to: Communications Department, Rand, 1700 Main Street, Santa Monica, California 90406.

Published by The Rand Corporation

PREFACE

This report is one of a series giving the results of experiments in group judgment. Previous experiments have been reported in references [1, 2, 3, 4]. The primary goal of these studies is the design of improved techniques for the use of expert opinion by decisionmakers. For many basic military issues the best information available is the judgment of knowledgeable individuals. Thus the military has an important stake in ensuring that the procedures used for obtaining judgments are designed to elicit the best judgments possible from the community of experts.

In practice, the advice received from experts is of two sorts: one dealing with matters of fact and one dealing with evaluation (criteria, priorities, goals, objectives, and so forth). Both kinds are important with respect to making effective use of advisers for military decisions. Previous studies have shown that it is possible to design improved techniques for using group judgment concerning matters of fact. The present report describes experiments to assess the appropriateness of similar techniques applied to matters of evaluation. Since the subjects for the experiment were college students, the material dealt with is somewhat removed from military issues. But the results support the conclusion that Delphi procedures are appropriate (in a well-defined sense) for the formulation and assessment of criteria and objectives.

This research was conducted for the Advanced Research Projects Agency.

ACKNOWLEDGMENTS

The authors wish to express their appreciation for the help of Ralph J. Lewis, consultant to The Rand Corporation, in the design and execution of this study. They also would like to thank Samuel Cochran (now at the University of Southeastern Texas) for his participation in conducting the experiment.

SUMMARY

This Report describes the results of an experiment assessing the appropriateness of Delphi procedures for formulating group value judgments. Upperclass and graduate students from UCLA were paid to act as subjects. The task was to generate and rate value categories relating to higher education and the quality of life. Two groups of forty subjects each generated lists of value categories which they considered important in the two respective areas. The initial lists — 300 and 250 items respectively — were aggregated by the experimental team to 45 and 48 items respectively. The subjects then rated all possible pairs of these items with respect to their similarity. The average similarity ratings were analysed by a clustering routine. Fifteen clusters in education and thirteen in quality of life were identified. These clusters were rated by the subjects with respect to their relative importance, with four subgroups using different rating procedures. The rating procedure was iterated once, with feedback on the second round of the medians and quartiles of the first-round ratings. The subjects from both groups then made estimates of the relative contribution of each of the educational categories to each of the quality of life categories.

The primary data analyses concerned the importance ratings. Three aspects were examined: (1) The quality of the distributions of the responses, (2) The correlation

between ratings by different groups and different rating techniques, and (3) The amount of change and degree of convergence upon iteration with feedback. As expected, the analyses showed that the distributions were in almost all cases single peaked and roughly bell-shaped; the correlations between both different groups and different rating methods were high (in the nineties); the number of changes and degree of convergence (reduction in standard deviation) were comparable to similar indices for factual judgments. The experiment furnished support for the conclusion that Delphi procedures are appropriate for processing value material as well as factual material.

Although the experiment was primarily concerned with assessing the use of Delphi techniques for processing value judgments, the substantive data appears to be of some interest on its own as an exploratory investigation of objectives for higher education and individual life. For example, a reweighting of the educational factors in terms of their summed contribution to the quality of life categories was compiled. The reweighted assessments showed large differences from the direct ratings, indicating the possibility that current notions of the role of the university are somewhat loosely tied to the basic interests of the students.

CONTENTS

PREFACE iii

ACKNOWLEDGMENTS v

SUMMARY vii

Section
1. INTRODUCTION 1

2. METHOD 8
2.1 Subjects 9
2.2 Item Generation 10
2.3 Importance Rating 18

3. RESULTS 25

4. DISCUSSION 42

REFERENCES 49

GROUP VALUE JUDGMENTS

1. INTRODUCTION

The last few years have seen a rapid increase in applications of group judgment techniques to public and corporate policymaking. One of the more widely applied techniques is Delphi, a term referring to a more or less specific set of procedures developed at The Rand Corporation for eliciting and processing the opinions of a group [1,2,3,4]. A rather extensive set of experiments has demonstrated that for subject matters where the best available information is the judgments of knowledgeable individuals, a systematic and controlled process of querying and aggregating the judgments of members of a group has distinct advantages over the traditional group discussion [1].

Most of the experiments which have been conducted to date have dealt with factual material. However, in some applications, the procedures have been employed to deal with a quite different sort of material, namely, value judgments. Typical is the use of Delphi procedures to identify and rate the objectives of industrial enterprises or to assess the relative importance of military missions. From the standpoint of the decisionmaker, opinions about values and objectives are just as relevant to decisions as factual opinions about consequences. Hence, the question whether Delphi procedures demonstrate advantages with value material of the same sort as those for factual material is a question of direct importance.

There are a number of difficulties in attempting to conduct experiments dealing with the excellence of value judgments. Above all, there is no generally agreed-upon way to measure the correctness of such judgments. Although there is some disagreement with respect to the proper measure for predictions of future events,* it is generally agreed that one relevant measure of excellence for factual opinions is just how close those opinions are to the true state of affairs. In general it is not difficult to arrange some scale whereby "closeness to the state of affairs" can be measured; although for opinions about the future, the investigator may have to bide his time until the future evolves. But in the case of value judgments, there is no generally agreed-upon corpus of "facts" against which the judgments can be compared.

Another difficulty with assessing the quality of value judgments has often been alleged: that they are "emotionally loaded." Expression of such judgments is more directly tied to emotions than factual statements; furthermore, commitment to those judgments is more central to the personality of the individual, so that the interaction of value judgments and other cognitive material is impeded [5].

These difficulties might be considered enough to discourage any "objective" measurement of the excellence of

---

*De Jouvenel [6] refers to futuribles as something different from states of affairs past and present; and some writers have been concerned about self-defeating or self-validating predictions [7].

such judgments. There is, of course, one type of objective study where there is no particular difficulty: that in which value judgments are considered simply as one aspect of human behavior, with no direct concern with what the judgments are about. Thus it is possible to study the genesis of judgments, the interrelationships between value systems, etc., without ever exploring the subject matter of these and especially without asking whether they are good or bad judgments.

However, this point of view is not the concern of the present inquiry. The usual point of view is that value judgments can be, in some sense, right or wrong. For example, when a corporate entity, e.g., a board of directors of an industrial firm, asks what are the objectives of their organization, what are their priorities, which objectives are crucial and which only desirable, it appears fairly clear that they are not asking, "What are our capricious feelings about what we should do?" They would not be willing to accept the assertion that any other set of whimsical attitudes would be just as reasonable as the ones they express.

The same is true of the values people express with regard to everyday life, or the set of values that are ascribed to the nation. There may be violent disagreements on all of these, but there is little disagreement that the judgments themselves are usually <u>not</u> capricious.

It appears, then, to be the case in disagreements about values that most individuals would state that one side can be more correct and the other less correct without being able to specify how the value judgments can be validated. Exceptions are usually referred to as "matters of taste." As it turns out, it is not necessary to be able to specify what correctness or incorrectness means in order to say a great deal about better and worse judgments.

If a group of indistinguishable experts expresses a range of opinions concerning a given question, then the median opinion of the group is more likely to be correct* than that of an (unspecified) member of the group [1, p. 7]. In a like manner, if a group of equally competent individuals expresses a range of opinions concerning a value question, then the average opinion is more likely to approximate the correct answer than an individual judgment, given the presumption that there is a correct answer to the value question. In order to make this assertion logically acceptable it is necessary to assume that the value judgment can be expressed in numerical terms. It appears that in most cases of practical import this can be done.

There are some other useful tautological consequences of the assumption that there is a correct answer to a value question. One is that the larger the group (maintaining

---

*Strictly speaking, this should be read "at least as likely to be correct."

indistinguishability), the more accurate the answer on the average. The other is that the larger the group, the greater the reliability of the answer, that is, the higher the probability that a similar group will express a similar answer.

All of these favorable aspects of group value judgments depend in part upon the degree to which it is considered that the group is judging something rather than simply reporting personal attitudes. Since we are precluded at the present time from a direct comparison of the group responses and an objective criterion, something weaker in the way of assessing the judgments is required. This something weaker is furnished by considering three of the necessary (but not sufficient) conditions for assuming there is a group judgment involved: These three conditions could be interpreted as a partial definition of the term group judgment for value questions.

(1) Reasonable distributions. If the distribution of group responses on a given numerical value judgment is flat, indicating group indifference, or if it is U-shaped, indicating either that the question is being interpreted differently by two subgroups, or actual difference of assessment by two subgroups, then it seems inappropriate to assert that the group considered as a unit has a judgment on that question.

(2) Group reliability. Given two similar groups (e.g., two groups selected out of a larger group at random) the group judgments on a given value question should be similar. Over a set of such value judgments, the correlation

for the two subgroups should be high.

(3) Change, and convergence on iteration with feedback. This condition is proposed in part by analogy with results from experiments with factual material, that is, shifts of individual responses toward the group response and reduction in group variability. More generally, if members of the group do not utilize the information in reports of the group response on earlier rounds when generating responses on later rounds, it seems inappropriate to consider these responses as judgments.

In the experiments described below, these three criteria are applied to value judgments by university students concerning the objectives of a higher education and the objectives of everyday (individual) life. The students generated a list of objectives for these two areas, and rated them on a scale of relative importance. Three different rating methods were employed in order to test both group reliability and stability over scaling technique. Ratings were obtained on each of two rounds, where the results of the first round (the median and upper and lower quartiles of the responses) were fed back between rounds. The data generated by the value judgments satisfied the three criteria to about the same degree as corresponding data from similar groups making factual estimations. In short, the outcome of these experiments appears to be that the Delphi procedures — as far as we can evaluate them at present — are appropriate for generating and assessing value material.

The primary purpose of the experiments was to evaluate the Delphi procedures for value material, but the data generated concerning what the subjects considered important with respect to a higher education and to everyday life appears to have some interest in its own right. This aspect of the experiment will be discussed more fully in the final section of this report.

2. METHOD

In this study one group of subjects used the Delphi procedure to rate the relative importance of each of a set of factors in terms of the factor's contribution to a person's assessment of the "Quality of Life." (In our instructions to the subjects we defined the term "Quality of Life" (QOL) to mean a person's sense of well being, his satisfaction or dissatisfaction with life, or his happiness or unhappiness.) A second group used the Delphi method to scale a set of changes in characteristics of students occurring as a result of their participation in the process of higher education. This scale measured the Effects of Education (EE) in terms of the importance of the changes for the student. These topics were selected because our subject population (UCLA upper-division and graduate students) could be expected to have informed opinions concerning each of them. The two groups received nearly the same instructions for the different topics and were for the most part treated identically.

The experiment required three sessions, the first two of which were devoted to the generation of the items to be scaled by the Delphi method in the third session. In the first session, each subject made up a list of from 5 to 10 items important either for the assessment of the Quality of Life or for the evaluation of the Effects of Education on students.

The items from the QOL group (about 250 in all) were sorted into 48 categories of similar items, while

the 300 items from the EE group were sorted into 45 categories. In the second session of the experiment the subjects who had made up the lists of items in response to the QOL questionnaire rated the similarity of all possible pairs of categories formed from the original QOL items. The EE group rated the similarity of all pairs of the EE categories. The similarity ratings were used to cluster the categories of the original items into super-categories. Thirteen super-categories or factors were formed from the QOL categories and fifteen from the EE categories. The relative importance of each factor was assessed during the third session of the study. The QOL group rated the importance of the QOL factors and the EE group rated the EE factors. A two-round Delphi procedure was employed where both groups revised their importance ratings during the second round in view of the median ratings for each factor obtained from the group's first-round ratings. As a check on the reliability of the ratings, the QOL and EE groups were each split into two subgroups and each subgroup used a different procedure to scale the factors.

2.1. Subjects

The subjects were 90 UCLA upper-division and graduate students. They were recruited by advertisements in the school paper and were paid for their participation. No attempt was made to select subjects according to sex or field of interest.

2.2. Item Generation

During the first session, which was conducted at UCLA, subjects were instructed to list from 5 to 10 items pertaining either to the "Quality of Life" or the "Effects of Education." The subjects were randomly assigned to a particular topic so that 45 subjects responded to each.

Subjects in the two groups were treated identically. The subjects were given printed instructions and a deck of 10 blank cards. The instructions briefly introduced the subject to the purpose of the experiment and then requested him to list from 5 to 10 items (one item per card) pertaining either to the QOL or the EE topic.

In the QOL condition, subjects were asked to list the characteristics or attributes of those events having the strongest influence on determining the QOL of an adult American. The subjects were instructed to ignore events concerned with basic biological maintenance, but not to overlook characteristics with negative connotations, e.g., aggression. Subjects in the EE condition were asked to view higher education as a process which causes (or fails to cause) changes in characteristics of students. The subjects were requested to list those characteristics which should be considered in evaluating the process of higher education. Subjects were instructed to consider only undergraduate education while forming their lists.

Subjects were also instructed to rank their items from most important to least important. These ranks were used only as rough guides in the initial aggregation of items by the experimental team. Questions concerning the experiment were answered either by repeating or paraphrasing the instructions. No subject required more than half an hour to complete the first session. They were then given appointments for the second and third sessions which were conducted at The Rand Corporation in Santa Monica at intervals of one week.

Prior to the second session of the experiment, the items generated by the subjects in the first session were sorted into categories of similar items. Two sets of categories were formed: one for the QOL items and another for the EE items. The sorting was done by a panel of three; each member assisted in the design and execution of the experiment. Two criteria were used during the sorting of the items: (1) The perceived differences of any pair of items within a category were to be smaller than differences between any pair of items drawn from two different categories; and (2) No more than 50 categories were to be formed. Composite labels were developed for each category either by quoting or paraphrasing (or both) a few of the most frequently occurring items in each of the categories. The 48 QOL category composite labels are given in Table 1 and the 45 EE composite labels are shown in Table 2.

Table 1

CHARACTERISTICS OF QUALITY OF LIFE

1. FEAR, ANXIETY
2. AGGRESSION, VIOLENCE, HOSTILITY
3. AMBITION
4. COMPETITION, COMPETITIVENESS
5. OPPORTUNITY, SOCIAL MOBILITY, LUCK
6. DOMINANCE, SUPERIORITY
7. MONEY, ACQUISITIVENESS, MATERIAL GREED
8. COMFORT, ECONOMIC WELL-BEING
9. NOVELTY, CHANGE, NEWNESS, VARIETY, SURPRISE
10. HONESTY, SINCERITY, TRUTHFULNESS
11. TOLERANCE, ACCEPTANCE OF OTHERS
12. STATUS, REPUTATION, RECOGNITION, PRESTIGE
13. FLATTERY, POSITIVE FEEDBACK, REINFORCEMENT
14. SPONTANEITY, IMPULSIVE, UNINHIBITED
15. FREEDOM
16. COMMUNICATION, INTERPERSONAL UNDERSTANDING
17. LONELINESS, IMPERSONALITY
18. DEPENDENCE, IMPOTENCE, HELPLESSNESS
19. POWER, CONTROL, INDEPENDENCE
20. GOOD HEALTH
21. FAILURE, DEFEAT, LOSING
22. INVOLVEMENT, PARTICIPATION
23. LOVE, CARING, AFFECTION
24. SELF-RESPECT, SELF-ACCEPTANCE, SELF-SATISFACTION
25. SELF-KNOWLEDGE, SELF-AWARENESS, GROWTH
26. SELF-CONFIDENCE, EGOISM
27. SECURITY
28. CHALLENGE, STIMULATION
29. PRIVACY
30. BOREDOM
31. ESCAPE, FANTASY
32. CONCERN, ALTRUISM, CONSIDERATION
33. HUMOR, AMUSING, WITTY
34. RELAXATION, LEISURE
35. SEX, SEXUAL SATISFACTION, SEXUAL PLEASURE
36. SUCCESS
37. ACHIEVEMENT, ACCOMPLISHMENT, JOB SATISFACTION
38. FAITH, RELIGIOUS AWARENESS
39. PEACE OF MIND, EMOTIONAL STABILITY, LACK OF CONFLICT
40. SUFFERING, PAIN
41. STABILITY, FAMILIARITY, SENSE OF PERMANENCE
42. INDIVIDUALITY
43. HUMILIATION, BELITTLEMENT
44. BEING NEEDED, FEELING OF BEING WANTED
45. CONFORMITY
46. SOCIAL ACCEPTANCE, POPULARITY
47. FRIENDSHIP, COMPANIONSHIP
48. EDUCATIONAL, INTELLECTUALLY STIMULATING

Table 2

CHARACTERISTICS OF EFFECTIVENESS OF HIGHER EDUCATION

1. SELF-AWARENESS, INCREASED SELF-UNDERSTANDING
2. MATURITY
3. ABILITY TO LEARN, LEARNING TO LEARN
4. CRITICAL ABILITY, QUESTIONING, DEVELOPMENT OF A CRITICAL ATTITUDE
5. HONESTY, PERSONAL INTEGRITY
6. CURIOSITY, DESIRE TO LEARN MORE
7. SOCIAL AWARENESS, AWARENESS OF OTHERS
8. SOCIAL CONTACTS, OPPORTUNITY TO MEET A VARIETY OF PEOPLE
9. TOLERANCE, DECREASE IN PREJUDICES
10. OPEN-MINDEDNESS
11. UNDERSTANDING OF OTHERS
12. CULTURAL AWARENESS
13. SOCIAL ISSUES, AWARENESS OF SOCIETAL PROBLEMS
14. SOCIAL SKILLS, ABILITY TO GET ALONG WITH OTHERS
15. BROADER OUTLOOK, NEW PERSPECTIVES, SCOPE
16. POLITICAL MATURITY, POLITICAL AWARENESS
17. COMMUNICATION SKILLS
18. KNOWLEDGE
19. DEHUMANIZATION, REPRESSIVE BUREAUCRACY
20. CAREER SKILLS, JOB COMPETENCE
21. SPECIALIZATION, NARROWING OF INTEREST TO OWN FIELD
22. REASONING ABILITIES, ABILITY TO THINK
23. SELF-CONFIDENCE, SELF-RELIANCE, INDEPENDENCE
24. IRRELEVANCY, PRESCRIBED EDUCATION, EDUCATIONAL TRIVIA
25. MOTIVATION, COMPETITIVENESS
27. IMPRACTICAL EDUCATION, DISILLUSIONMENT WITH EDUCATIONAL USEFULNESS
28. LOSS OF CREATIVITY, LOSS OF CREATIVE THINKING
29. GREATER CREATIVITY, EXPANDING THE IMAGINATION
30. LOSS OF IDEALISM, GENERAL DISSATISFACTION
31. RESPONSIBILITY
32. SEXUAL MATURITY, MORE LIBERAL SEXUAL ATTITUDE
33. POLITICAL DISILLUSIONMENT
34. NEW EXPERIENCES, EXPOSING TO NEW ACTIVITIES
35. NARROWING OF OUTLOOK, NARROWING OF VALUES
36. SELF-RESPECT, SELF-ACCEPTANCE, SELF-SATISFACTION
37. DEPENDENCY, PROLONGED YOUTH
38. SYNTHESIZING ABILITY, A SENSE OF ORGANIC RELATIONSHIP
39. AWARENESS OF ENVIRONMENT, RELATIONSHIP OF INDIVIDUAL WITH ENVIRONMENT
40. LIBERALIZATION OF SOCIAL AND POLITICAL VIEWS
41. PURPOSE IN LIFE, DEVELOPMENT OF LIFE GOALS
42. ELITISM, SOCIAL STATUS
43. INVOLVEMENT, POLITICAL INVOLVEMENT
44. CONCERN FOR SOCIETY, FELLOWMAN
45. RELATING TO OTHERS

During the second session, each subject was presented with a list of all possible pairs of either the QOL or EE category labels. The task for all subjects was to rate the similarity of the labels in each pair. Every subject was given printed instructions, a list of the category labels, and a computer-generated list of pairs of labels. Each subject received a different random ordering of label pairs. The instructions informed the subjects that the items they had developed during the first session had been categorized to form the list of category labels. This list had in turn been used to form the computer printed list of label pairs. The subjects were instructed to rate the similarity of the labels in each pair on a 0-4 scale where the numerical ratings were tied to the following adjective scale:

4 Practically the same

3 Closely related

2 Moderately related

1 Slightly related

0 Unrelated

If a subject felt that the labels were connected, but in an inverse fashion, he was to use negative ratings, e.g., -4 being equivalent to "practically opposites." The following two examples were given; Drowsy - Physically Tired, illustratively scored at 2, and Drowsy - Alert, scored at -3. Both groups received the same instructions. The QOL group

rated 1128 item pairs and the EE group rated 990. The experiment was conducted in two 1-1/4-hour periods with a 1/2-hour break between periods.

The means of the absolute values of the similarity ratings for each label pair were computed over subjects for both groups. These mean absolute ratings were then analyzed by Johnson's hierarchical clustering procedure [8]. In this procedure objects are clustered according to the similarities between them. The objects within a cluster are more similar to one another than to objects belonging to a different cluster. In addition, the procedure merges similar clusters into larger clusters in a step-wise fashion until all the objects are placed into a single cluster. Consequently, the user of this procedure must select the number of clusters which seems compatible with both the data and any theoretical or empirical predictions about the results of the procedure. The problem is not unlike selecting the number of factors to retain in a factor analysis. The use of the absolute values of the ratings "folds" the label pairs given the negative ratings into the same clusters. The clusters which were generated by this procedure are shown in Fig. 1 for the QOL groups and Fig. 2 for the EE group. Numbers across the top refer to the list of items in Tables 1 and 2 respectively. The left-hand column indicates the similarity level at which the item is included in a cluster. The "histogram" of x's displays the progressive aggregation of items into clusters. For example, in

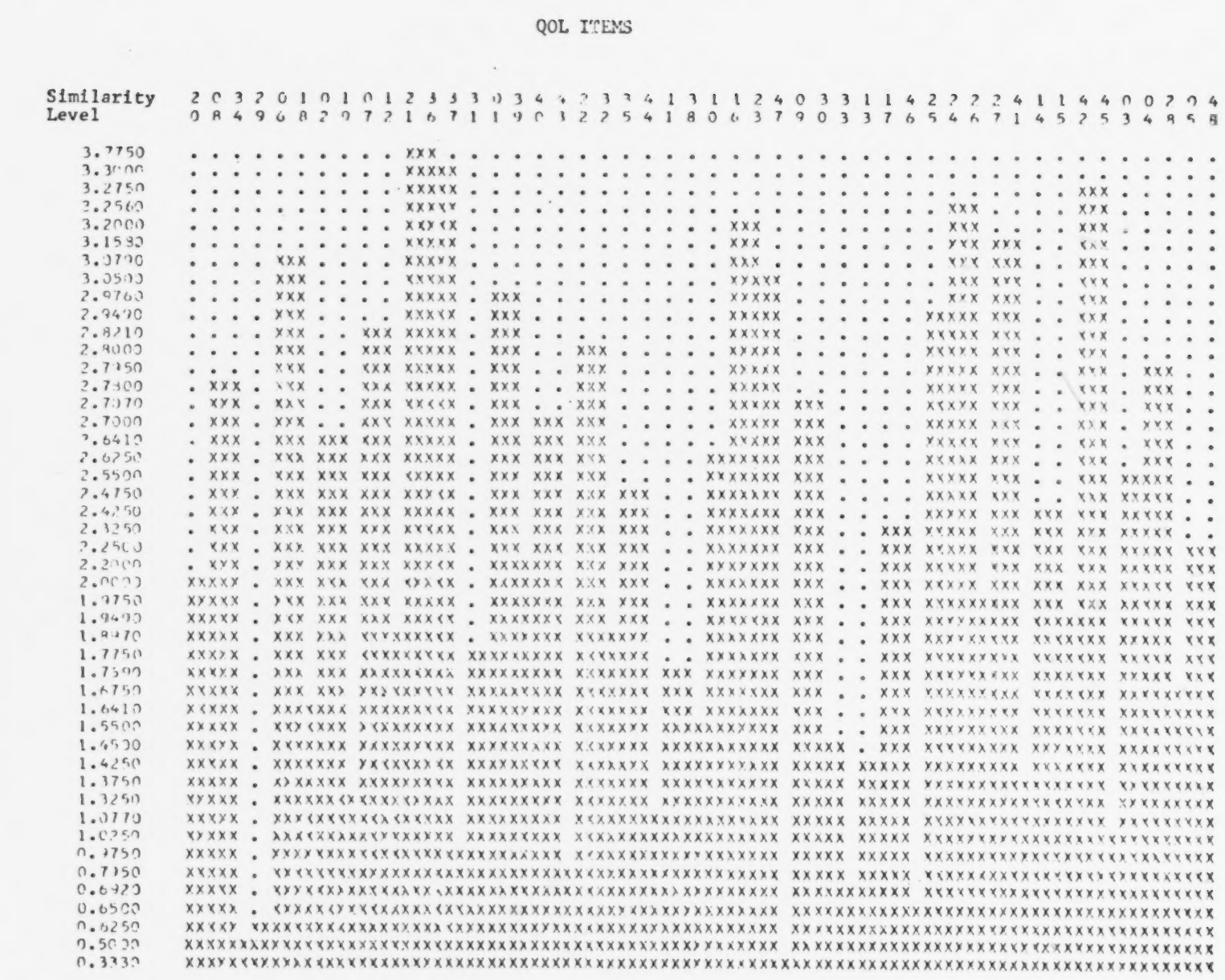

Fig. 1—Computer-generated display of factors from analysis of QOL similarity ratings

MEANS OF ABSOLUTE VALUES OF SIMILARITY RATINGS

EE ITEMS

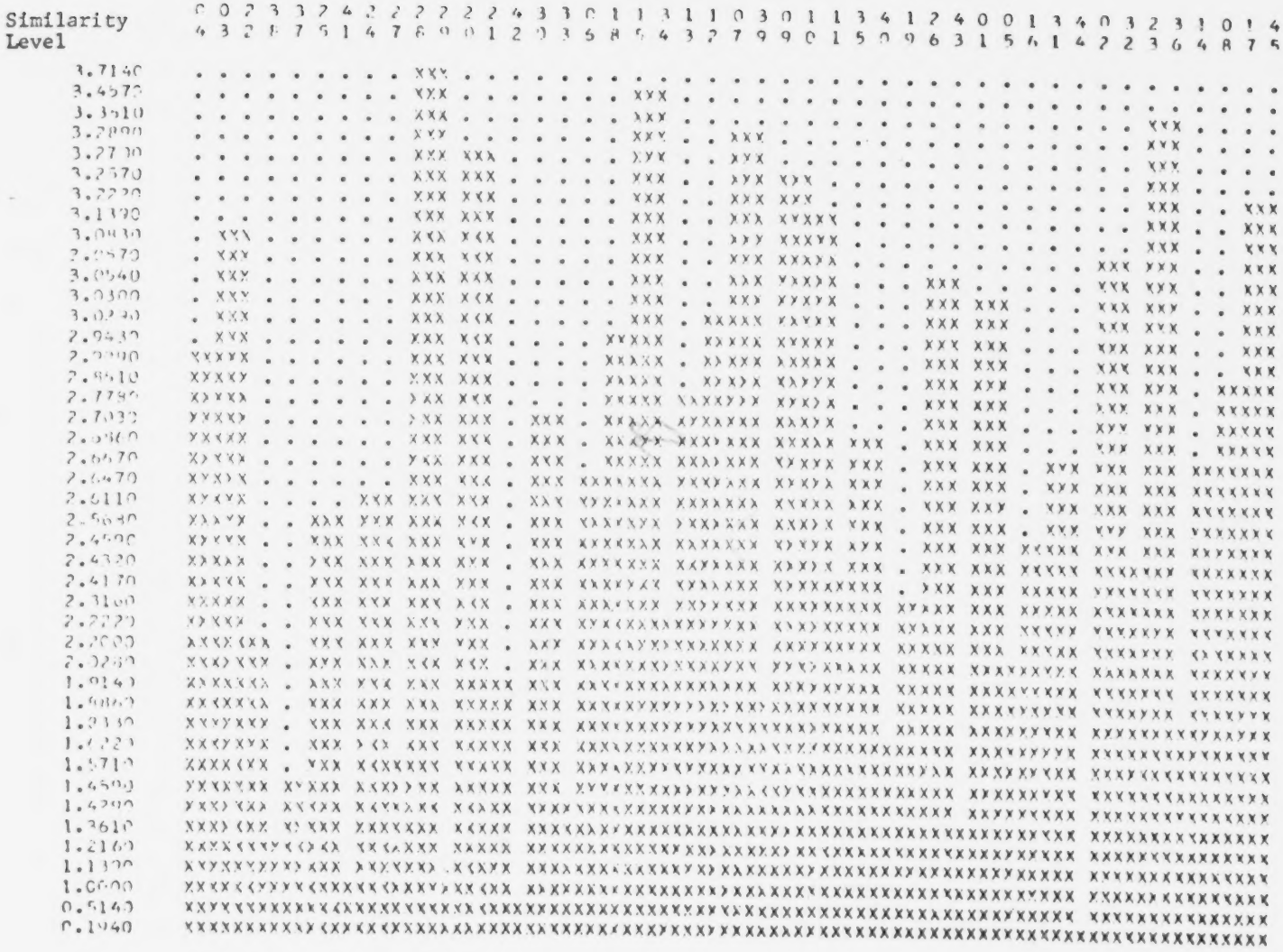

Fig. 1 at the highest level of similarity (3.78) Failure (21) and Success (35) are associated — probably as straightforward opposites. At almost the same level, Achievement (37) is joined to the cluster. Nothing further is added to this cluster until level 1.9 when the previously associated pair, Money (7) and Status (12) are added. This is the "core" of characteristic 11 in Table 3. The thirteen QOL and fifteen EE clusters which were selected are given in Tables 3 and 4.

2.3 Importance Rating

The task for the subjects in the third session of the experiment was to rate the clusters or factors in terms of their importance to the topic in question. The subjects who had developed the QOL factors rated them as did the subjects who generated the EE factors. The design of this session is shown schematically in Table 5. As can be seen in Table 5, the QOL and EE groups were each split into two subgroups, and each subgroup used a different scaling procedure. During the third part of the session, the QOL and EE group both rated the relevance of each of the EE factors in terms of its contribution to each of the QOL factors. Otherwise, the groups were treated identically.

In order to familiarize the subjects with the factors they would be rating, they were instructed to look over the factors and devise a convenient word or phrase label for each. The subjects were then asked to rate their self-confidence in working with each of the factors on a 1

Table 3
QOL FACTORS

1. Novelty, change, newness, variety, surprise; boredom; humorous, amusing, witty.

2. Peace of mind, emotional stability, lack of conflict; fear, anxiety; suffering, pain; humiliation, belittlement; escape, fantasy.

3. Social acceptance, popularity; needed, feeling of being wanted; loneliness, impersonality; flattering, positive feedback, reinforcement.

4. Comfort, economic well-being; relaxation, leisure; good health.

5. Dominance, superiority; dependence, impotence, helplessness; aggression, violence, hostility; power, control, independence.

6. Challenge, stimulation; competition, competitiveness; ambition; opportunity, social mobility, luck; educational, intellectually stimulating.

7. Self-respect, self-acceptence, self-satisfaction; self-confidence, egoism; security; stability, familiarity, sense of permanence; self-knowledge, self-awareness, growth.

8. Privacy.

9. Involvement, participation; concern, altruism, consideration.

10. Love, caring, affection; communication, interpersonal understanding; friendship, companionship; honesty, sincerity, truthfulness; tolerance, acceptance of others; faith, religious awareness.

11. Achievement, accomplishment, job satisfaction; success; failure, defeat, losing; money, acquisitiveness, material greed; status, reputation, recognition, prestige.

12. Individuality; conformity; spontaneity, impulsive, uninhibited; freedom.

13. Sex, sexual satisfaction, sexual pleasure.

Table 4

EDUCATIONAL FACTORS

1. Greater creativity, expanding the imagination; loss of creativity, loss of creative thinking.
2. Broader outlook, new perspectives, scope, new experiences, exposing to new activities; knowledge; curiosity, desire to learn more.
3. Social awareness, awareness of others; awareness of environment, relationship of individual with environment; cultural awareness; social issues, awareness of societal problems.
4. Career skills, job competence; specialization, narrowing of interest to own field; elitism, social status.
5. Involvement, political involvement; isolation from real world, ivory-tower syndrome; dehumanization, repressive bureaucracy.
6. Self-awareness, increased self-understanding; honesty, personal integrity.
7. Loss of idealism, general dissatisfaction; political disillusionment.
8. Self-confidence, self-reliance, independence; self-respect, self-acceptence, self-satisfaction; maturity; sexual maturity, more liberal sexual attitude.
9. Tolerance, decrease in prejudices; open-mindedness; understanding of others; narrowing of outlook, narrowing of values; liberalization of social and political views.
10. Communication skill; relating to others; social contacts, opportunity to meet a variety of people; social skills, ability to get along with others.
11. Responsibility; concern for society, fellowman; political maturity, political awareness.
12. Motivation, competitiveness, purpose in life, development of life goals.
13. Dependency, prolonged youth.
14. Ability to learn, learning to learn; reasoning abilities, ability to think, critical ability, questioning, development of a critical attitude; synthesizing ability, a sense of organic relationship.
15. Impractical education, disillusionment with educational usefulness; irrelevancy, prescribed education, educational trivia.

Table 5

STRUCTURE OF STUDENT JUDGMENTS FOR SESSION THREE

QOL Group		EE Group	
Subgroup 1 Split 100 N = 20	Subgroup 2 Magnitude Estimation N = 19	Subgroup 3 Split 100 N = 19	Subgroup 4 7-pt rating scale N = 18
Part 1			
Label factors	Label factors	Label factors	Label factors
Rate self-confidence with each factor on a 1—5 pt scale	Rate self-confidence with each factor on a 1—5 pt scale	Rate self-confidence with each factor on a 1—5 pt scale	Rate self-confidence with each factor on a 1—5 pt scale
Split 100 pts among the factors according to importance of each factor	Rate the most important factor with 100 pts and rate the other factors proportionately	Split 100 pts among the factors according to importance of each factor	Rate the importance of each factor on a 1 to 7 pt scale
Part 2			
Revise ratings in light of group median and quartiles for each factor	Revise ratings in light of group median and ranges for each factor	Revise ratings in light of group median and quartiles for each factor	Revise ratings in light of group median and quartiles for each factor
Part 3			
Rate the relevance of each EE factor to each of the QOL factors on a 0 to 3 point scale	Rate the relevance of each EE factor of each of the QOL factors on a 0 to 3 point scale	Rate the relevance of each EE factor to each of the QOL factors on a 0 to 3 point scale	Rate the relevance of each EE factor to each of the QOL factors on a 0 to 3 point scale

to 5-point scale. The factors they felt most confident about were to receive a 5 and those they felt least confident about were to receive a 1. Next the subjects were requested to rate the relative importance of each factor in terms of the contribution of that factor to the general topic. Using the split-100 (S-100) procedure, QOL Group 1 and EE Group 1 were instructed to distribute 100 points among the factors so that the most important factors received the most points. Using the magnitude-estimation (M-E) procedure QOL Group 2 was instructed to find the most important factor and give it a rating of 100. Then this group was asked to rate the other factors in terms of the most important one, so that a factor which they felt was half as important as the most important was to receive a rating of 50. The group using the rating scale (7-pt) procedure (EE Group 2) was asked to use a 1- to 7-point scale to rate the factors; a rating of 1 was to apply to "unimportant" factors, 4 to "moderately important" ones, and 7 to "extremely important" factors.

The subjects recorded their self-confidence ratings, factor labels, and importance ratings on preprinted response sheets. They also kept a record of their labels and importance ratings which they referred to during the second and third parts of the session.

During the second part of the session, the subjects agains rated the importance of the factors with the same method which they used during the first part. This time,

however, they were given information about the group's previous ratings on each of the factors. The QOL split-100, EE split-100, and EE 7-point rating scale groups were provided with the median and the first and third quartiles for each factor, while the QOL magnitude-estimation group was given ranges and medians which were normalized so that the largest median was 100. The instructions explained the meanings of the statistics and requested the subjects to consider this information in revising their estimates of the relative importance of each of the factors. The subjects were given 20 minutes to complete this part of the experiment.

During the third part of the session, the QOL and EE groups rated the "relevance" of each of the EE factors to each of the QOL factors. Each group received response sheets containing spaces along the top for each of the factor labels that they had developed during the first part of the session, and a list of QOL factors or EE factors, respectively, down the left margin. The subjects were briefly informed about the origin of the list of factors appearing on the left margin of their worksheets. Next, the subjects were instructed to familiarize themselves with these new lists of factors. Any questions concerning the list were answered by the experimenter. Finally, the subjects were required to rate the relevance of each of the EE factors to each of the QOL factors on a 0- to 3-point rating scale. Relevance

was defined in the instructions as either "contributing to" or "means the same thing as." The 0- to 3-point scale was tied to the following adjectives:

3 Contributes strongly (or is pretty much the same)

2 Contributes moderately

1 Contributes slightly

0 Irrelevant

The subjects were allowed 30 minutes for the completion of this part of the session.

3. RESULTS

Summary statistics computed from the QOL split-100 and QOL magnitude-estimation ratings on both rounds are given in Table 6. Similar statistics for the EE group are shown in Table 7. Both tables show the mean and median ratings and the standard deviations (SD) of the ratings for each factor. The factor identification numbers are keyed to the lists given in Table 3 for the QOL factors and Table 4 for the EE factors. In addition to the mean and median ratings, the geometric means (G-M) of the ratings are given for the QOL magnitude-estimation group. This was done in accordance with recommendations by Stevens [9] concerning the proper method of averaging magnitude estimates. Furthermore, the means, geometric means, and medians have been normalized so that the largest average rating is 100. These statistics are based on 20 cases for the QOL-100 group, 19 cases for the QOL magnitude-estimation and EE split-100 groups, and 18 cases for the EE 7-point rating scale group. The QOL factors are listed according to the decreasing split-100 second-round median ratings in Table 8. The EE factors are similarly listed in Table 9.

The agreement between the first-and second-round average ratings is very high for all four groups. The product-moment correlations between the median ratings on the first and second rounds is 0.99 for the QOL split-100 group, 0.97 for the QOL magnitude-estimation group, 0.97 for

Table 6

SUMMARY OF RESULTS FOR QOL GROUPS

Factor	Split 100, Round 1 N = 20			Split 100, Round 2 N = 20			Magnitude Estimation, Round 1 N = 19				Magnitude Estimation, Round 2 N = 19			
	Mean	Median	SD	Mean	Median	SD	Mean	G. Mean	Median	SD	Mean	G. Mean	Median	SD
1. Novelty	6.1	5.5	4.3	5.2	5.0	3.4	60	49	61	23	63	61	63	16
2. Peace of mind	9.2	10.0	3.6	9.3	10.0	3.1	94	94	91	16	96	96	94	10
3. Social acceptance	8.1	8.0	4.1	8.3	8.0	3.3	77	71	75	24	81	75	81	21
4. Comfort	6.4	6.5	2.9	6.1	6.0	2.3	72	69	63	22	73	71	71	19
5. Dominance	4.5	3.5	4.0	4.7	3.5	4.8	60	40	58	32	63	53	56	28
6. Challenge	8.3	8.0	4.8	8.0	8.0	3.3	72	65	80	25	80	76	82	18
7. Self-respect	11.1	10.5	4.4	11.9	11.5	4.1	98	98	100	15	99	99	100	13
8. Privacy	3.7	2.5	3.0	3.4	2.0	2.8	57	48	55	26	59	53	61	24
9. Involvement	6.7	6.5	3.4	6.0	6.0	2.6	72	69	72	22	73	71	74	18
10. Love	14.8	14.0	4.4	15.5	15.0	3.8	100	100	96	14	100	100	98	9
11. Achievement	7.6	7.5	4.6	7.5	7.0	4.2	79	78	79	16	79	78	78	15
12. Individuality	6.0	5.5	3.4	6.1	6.0	3.1	80	75	76	21	75	72	79	19
13. Sex	7.9	9.0	3.2	8.3	9.5	3.4	80	78	78	20	80	79	77	14

Table 7

SUMMARY OF RESULTS FOR EE GROUPS

Factor	Split 100, Round 1 N = 19			Split 100, Round 2 N = 19			7-point Rating Scale, Round 1 N = 18			7-point Rating Scale, Round 2 N = 18		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
1. Creativity	7.6	8.0	3.6	7.9	8.0	2.9	6.3	7.0	1.1	6.5	7.0	0.8
2. Broader outlook	11.3	10.0	5.0	11.4	10.0	3.5	6.6	7.0	0.7	6.7	7.0	0.5
3. Social awareness	8.1	7.0	4.5	8.3	8.0	3.8	6.1	6.5	1.1	6.3	6.5	0.9
4. Career skills	6.8	5.0	5.4	6.1	5.0	4.0	4.3	4.5	2.3	4.0	4.0	1.9
5. Involvement	3.9	5.0	2.2	4.8	5.0	2.7	3.9	3.5	1.7	3.4	3.0	1.5
6. Self-awareness	6.1	5.0	3.5	6.4	6.0	3.7	5.2	5.5	1.7	5.3	5.0	1.3
7. Loss of idealism	3.3	3.0	3.1	2.2	1.0	2.4	3.0	2.5	1.6	2.6	2.0	1.5
8. Self-confidence	6.3	6.0	3.2	6.6	6.0	2.2	5.4	5.0	1.5	5.3	5.0	1.1
9. Tolerance	6.6	5.0	4.3	6.8	6.0	3.6	4.9	5.5	1.7	5.2	5.5	1.3
10. Communication skill	6.7	6.0	3.7	6.8	7.0	3.2	4.8	6.0	2.0	4.9	6.0	1.7
11. Responsibility	4.6	5.0	2.0	4.8	5.0	2.1	5.1	5.0	1.8	4.9	5.0	1.6
12. Motivation	5.3	5.0	3.0	4.7	5.0	2.2	4.3	4.0	2.1	4.2	4.0	1.8
13. Dependence	2.2	1.0	2.6	1.2	0.0	1.7	2.7	1.5	2.2	2.1	1.0	1.5
14. Ability to learn	16.8	12.0	13.8	17.4	12.0	12.1	6.2	7.0	1.3	6.4	7.0	1.2
15. Impractical education	4.9	5.0	5.3	4.8	5.0	4.0	4.1	4.5	2.7	4.4	5.0	2.2

Table 8

	QOL FACTORS	Relative Importance
1.	Love, caring, affection, communication, interpersonal understanding; friendship, companionship; honesty, sincerity, truthfulness; tolerance, acceptance of others; faith, religious awareness.	15.0
2.	Self-respect, self-acceptance, self-satisfaction; self-confidence, egoism; security; stability, familiarity, sense of permanence; self-knowledge, self-awareness, growth.	11.5
3.	Peace of mind, emotional stability, lack of conflict; fear, anxiety; suffering, pain; humiliation, belittlement; escape, fantasy.	10.0
4.	Sex, sexual satisfaction, sexual pleasure.	9.5
5.	Challenge, stimulation; competition, competitiveness; ambition; opportunity, social mobility, luck; educational, intellectual stimulating.	8.0
6.	Social acceptance, popularity; needed, feeling of being wanted; loneliness, impersonality; flattering, positive feedback, reinforcement.	8.0
7.	Achievement, accomplishment, job satisfaction; success; failure, defeat, losing; money, acquisitiveness, material greed; status, reputation, recognition, prestige.	7.0
8.	Individuality; conformity; spontaneity, impulsive, uninhibited; freedom.	6.0
9.	Involvement, participation; concern, altruism, consideration.	6.0
10.	Comfort, economic well-being, relaxation, leisure; good health.	6.0
11.	Novelty, change, newness, variety, surprise; boredom; humorous, amusing, witty.	5.0
12.	Dominance, superiority; dependence, impotence, helplessness; aggression, violence, hostility; power, control, independence.	3.5
13.	Privacy.	2.0

Table 9

EE FACTORS

		Relative Importance
1.	Ability to learn, learning to learn; reasoning abilities, ability to think; critical ability, questioning, development of a critical attitude; synthesizing ability, a sense of organic relationship.	12.0
2.	Broader outlook, new perspectives, scope, new experiences, exposing to new activities; knowledge; curiosity, desire to learn more.	10.0
3.	Greater creativity, expanding the imagination, loss of creativity, loss of creative thinking.	8.0
4.	Social awareness, awareness of others; awareness of environment, relationship of individual with environment; cultural awareness; social issues, awareness of societal problems.	8.0
5.	Communication skill; relating to others; social contacts, opportunity to meet a variety of people; social skills, ability to get along with others.	7.0
6.	Tolerance, decrease in prejudices; open-mindedness; understanding of others; narrowing of outlook, narrowing of values; liberalization of social and political views.	6.0
7.	Self-awareness, increased self-understanding; honesty, personal integrity.	6.0
8.	Self-confidence, self-reliance, independence; self-respect, self-acceptance, self-satisfaction; maturity, sexual maturity, more liberal sexual attitude.	6.0
9.	Responsibility; concern for society, fellowman; political maturity, political awareness.	5.0
10.	Impractical education, disillusionment with educational usefulness; irrelevancy, prescribed education, educational trivia.	5.0
11.	Career skills, job competence; specialization, narrowing of interest to own field; elitism, social status.	5.0
12.	Motivation, competitiveness; purpose in life, development of life goals.	5.0
13.	Involvement, political involvement; isolation from real world, ivory-tower syndrome; dehumanization, repressive bureaucracy.	5.0
14.	Loss of idealism, general dissatisfaction; political disillusionment.	1.0
15.	Dependency, prolonged youth.	0.0

the EE split-100 group, and 0.99 for the EE 7-point rating scale group. The agreement between the rating methods for a given set of factors (reliability) is also very high. The plot of median magnitude estimation as a function of median split-100 rating for all the QOL factors is shown in Fig. 3. The open circles refer to the first round and the filled circles to the second. A similar graph for the EE factors is shown in Fig. 4; median 7-point rating is plotted as a function of median split-100 rating. Here again, the results for the first and second rounds are shown as open and filled circles, respectively. The correlation between the median (S-100) ratings and median median magnitude-estimation ratings is 0.90 on the first round and 0.91 on the second for the QOL factors. The correlation between the median (S-100) and 7-point ratings for the EE factors is 0.88 on the first round and 0.93 on the second. Note that in both cases Round 2 reliability was slightly greater than Round 1 reliability.

The greatest change in group performance between rounds is the decrease in response variability from the first to the second round. Round 2 standard deviations (SD) are generally smaller than corresponding Round 1 standard deviations, as is shown in Tables 6 and 7. The statistical significance of this decrease was assessed by comparing the mean of the SDs on the first round to the mean of the SDs on the second with t-tests [10, p. 170].

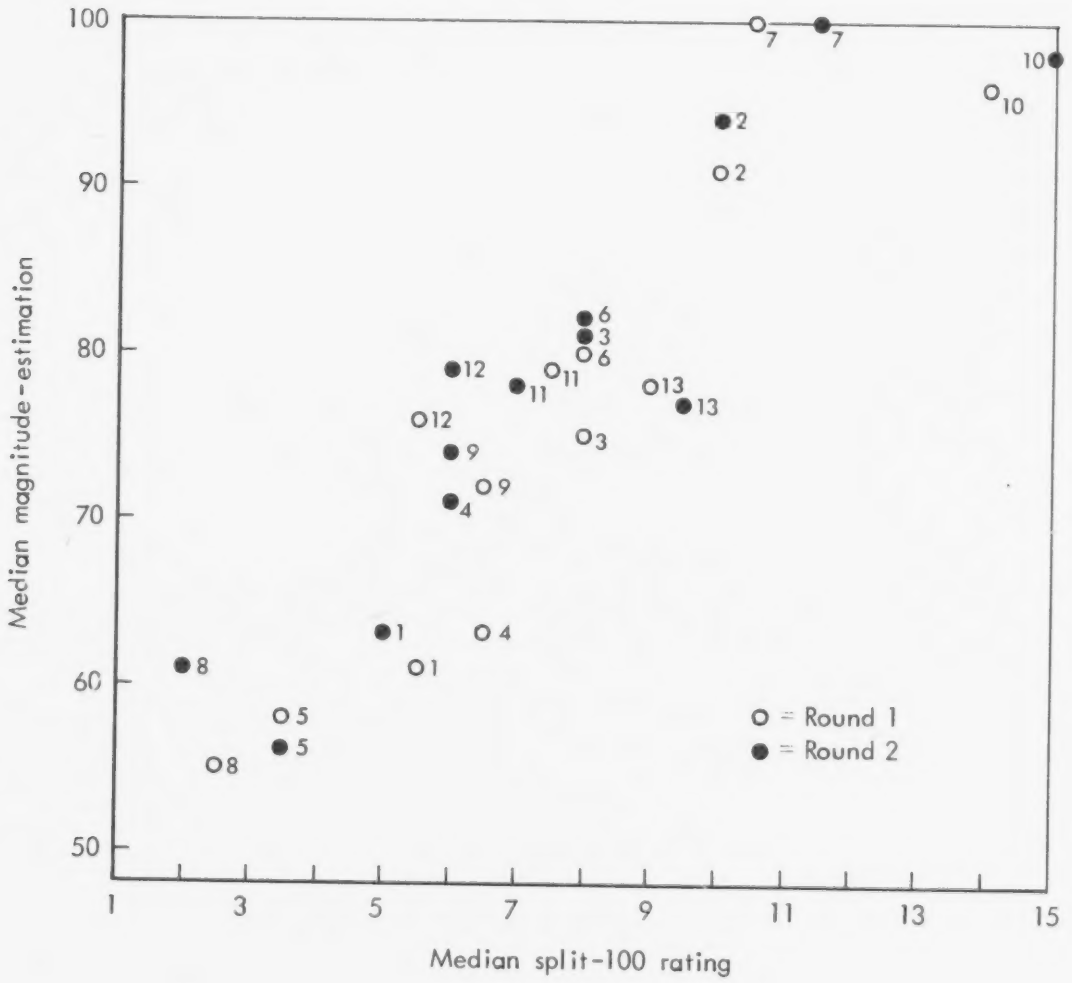

Fig.3—Cross-plot of split-100 and magnitude-estimation median ratings for QOL factors

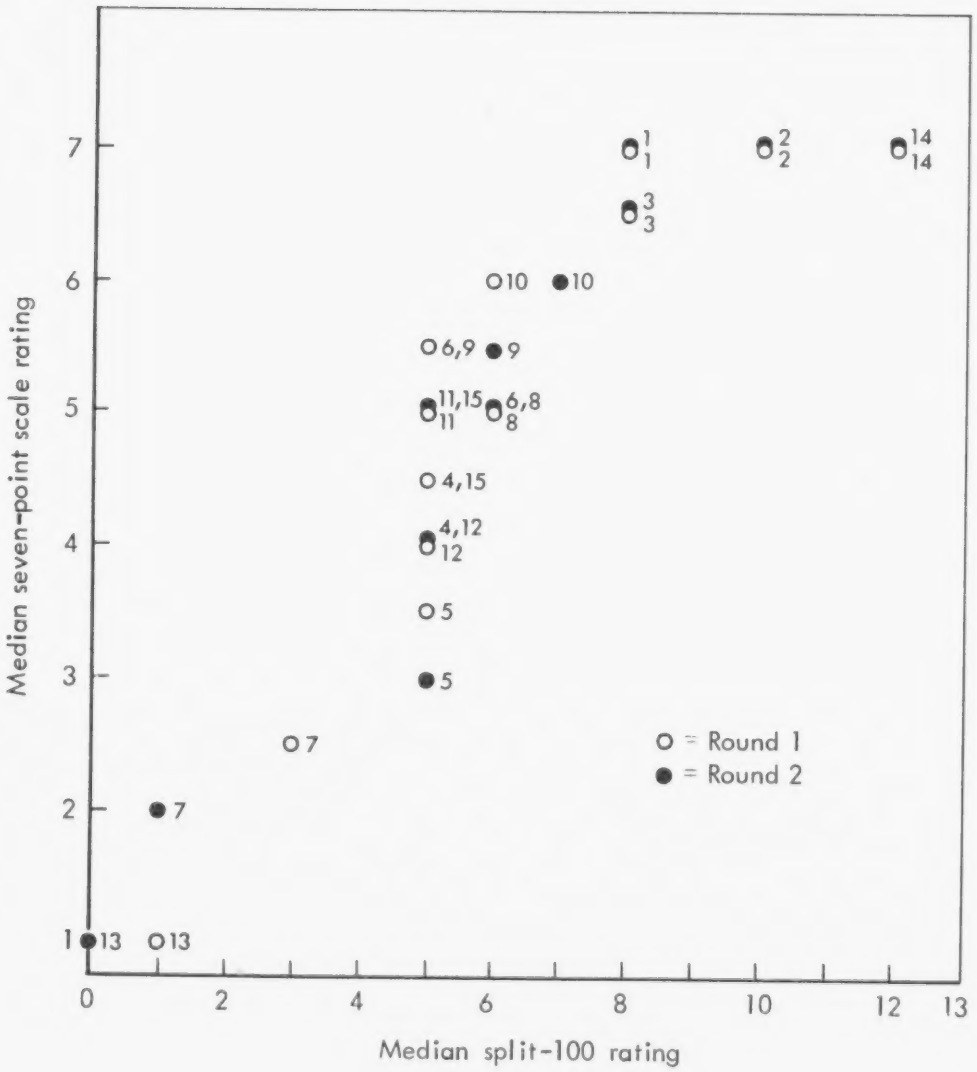

Fig.4—Cross-plot of split-100 and seven-point scale median ratings for EE factor

The mean SDs were computed over the factors. The mean differences were in the expected direction for all four groups. Round 1 and Round 2 differences are shown in Table 10. Computed $\underline{t}$'s and significance levels (p) are also shown. All differences were reliable at least at the 0.01 level.

Table 10

Differences Between Round 1 and Round 2 Mean Standard Deviations For All Groups

Item	QOL S-100	QOL ME	EE S-100	EE 7-Point
SD_1-SD_2	0.45	4.0	0.75	0.34
$\underline{t}$	3.03	7.34	4.58	7.59
df	12.00	12.00	14.00	14.00
p	<0.01	<0.005	<0.005	<0.005

The distributions of the responses to the questions in the previous Delphi experiments have been bell-shaped and generally positively skewed. In fact, the lognormal distribution has provided a very satisfactory fit to the observed distributions [1, p. 25]. These distributions of importance ratings were not fit to the lognormal distribution, but approximately equivalent bell-shaped distributions were expected for the ratings to each factor. In order to detect any deviant distributions the following procedure was

employed. First, the scores for each factor in each of the four groups were converted to deviation scores by subtracting the mean rating for a factor from each of the scores for the same factor. This centers the distributions of the ratings for all the factors about zero but does not change the variability, skewness, or kurtosis of the distributions. This transformed scale is used as the abscissa for Figs. 5, 6,7. Then the relative cumulative distribution for each factor was compared to the relative cumulative distribution for all the other factors combined in the same group and round with the Kolmogorov-Smirnov (K-S) two-sample test [10, p. 203]. The tests were made on both the first- and second-round ratings within each of the groups; altogether 112 tests were conducted. Only four distributions were found which differed from the composite distributions at the 10-percent significance level. The composite distributions are shown in Fig. 5 for the second-round ratings for the QOL Split-100, QOL magnitude-estimation, EE Split-100, and EE 7-point rating scale groups. The curves are all bell-shaped and generally skewed. The two most deviant distributions are shown in Fig. 6. Representative response distributions for the four groups on the second round are shown in Fig. 7. These were selected by choosing the response distribution within each group with the median p value according to the K-S tests.

Both groups of subjects (QOL and EE) rated the relevance of each EE factor to each of the QOL factors on a 0- to 3-

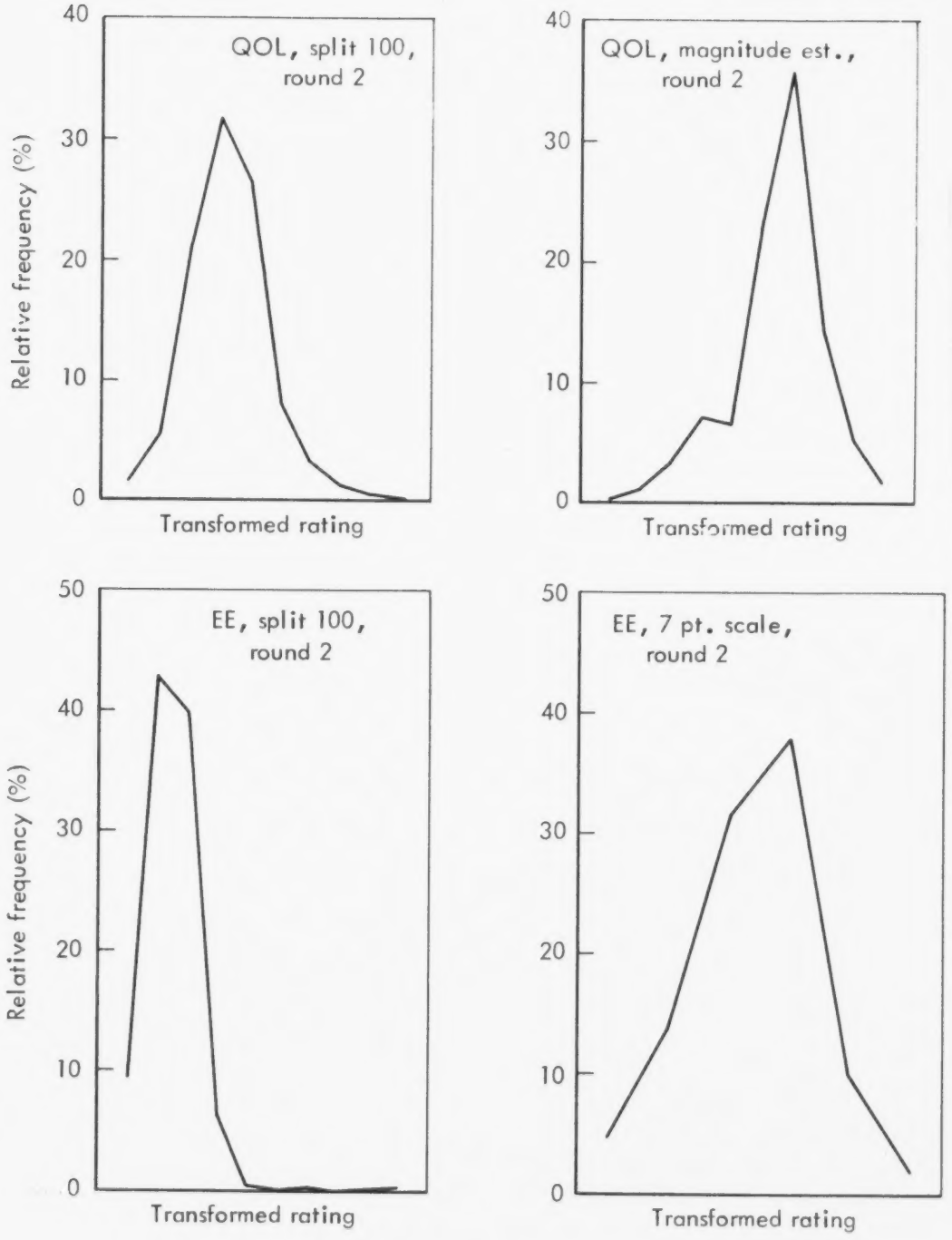

Fig.5—Average frequency distributions

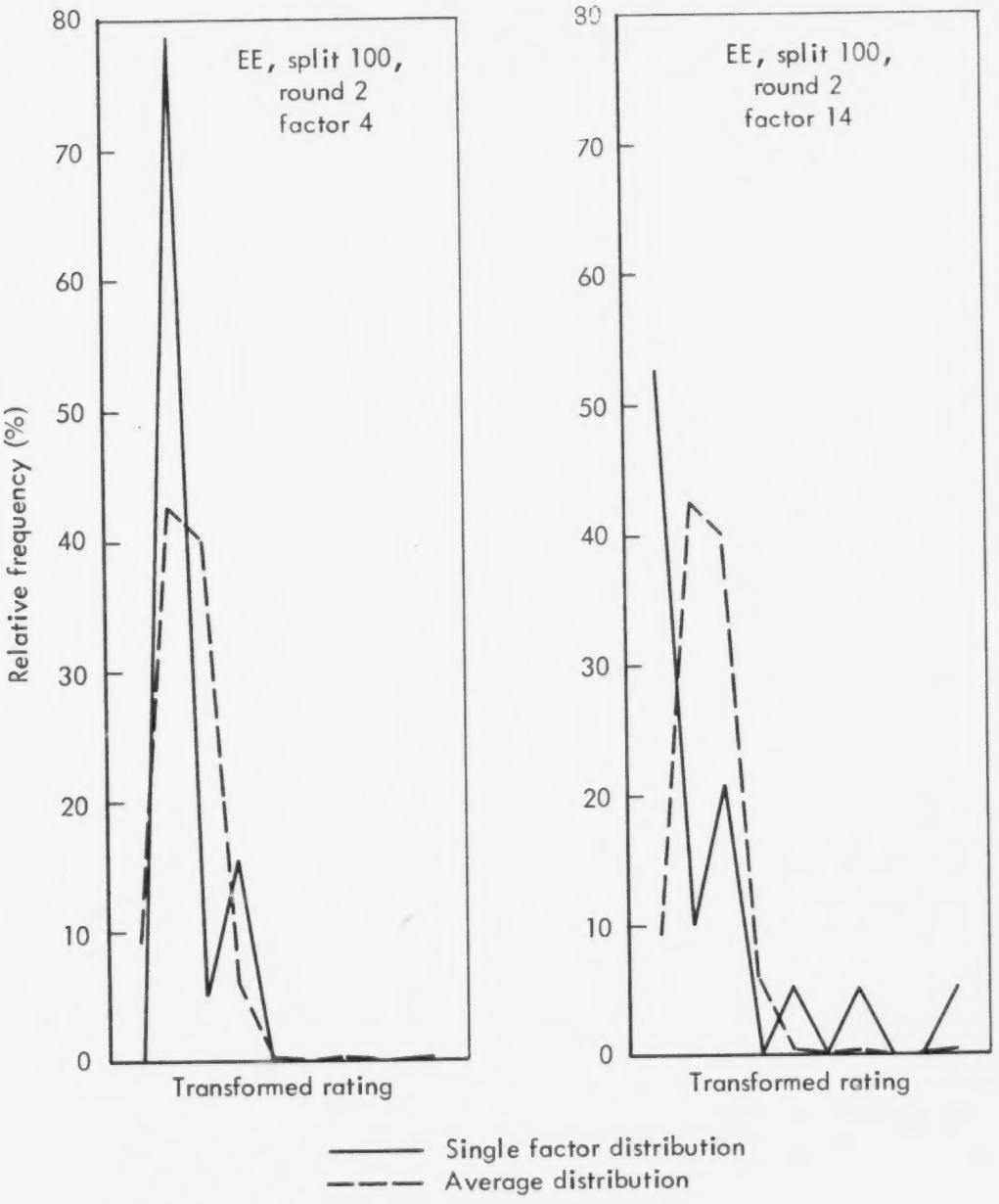

Fig.6—Most deviant distributions (for single factors)

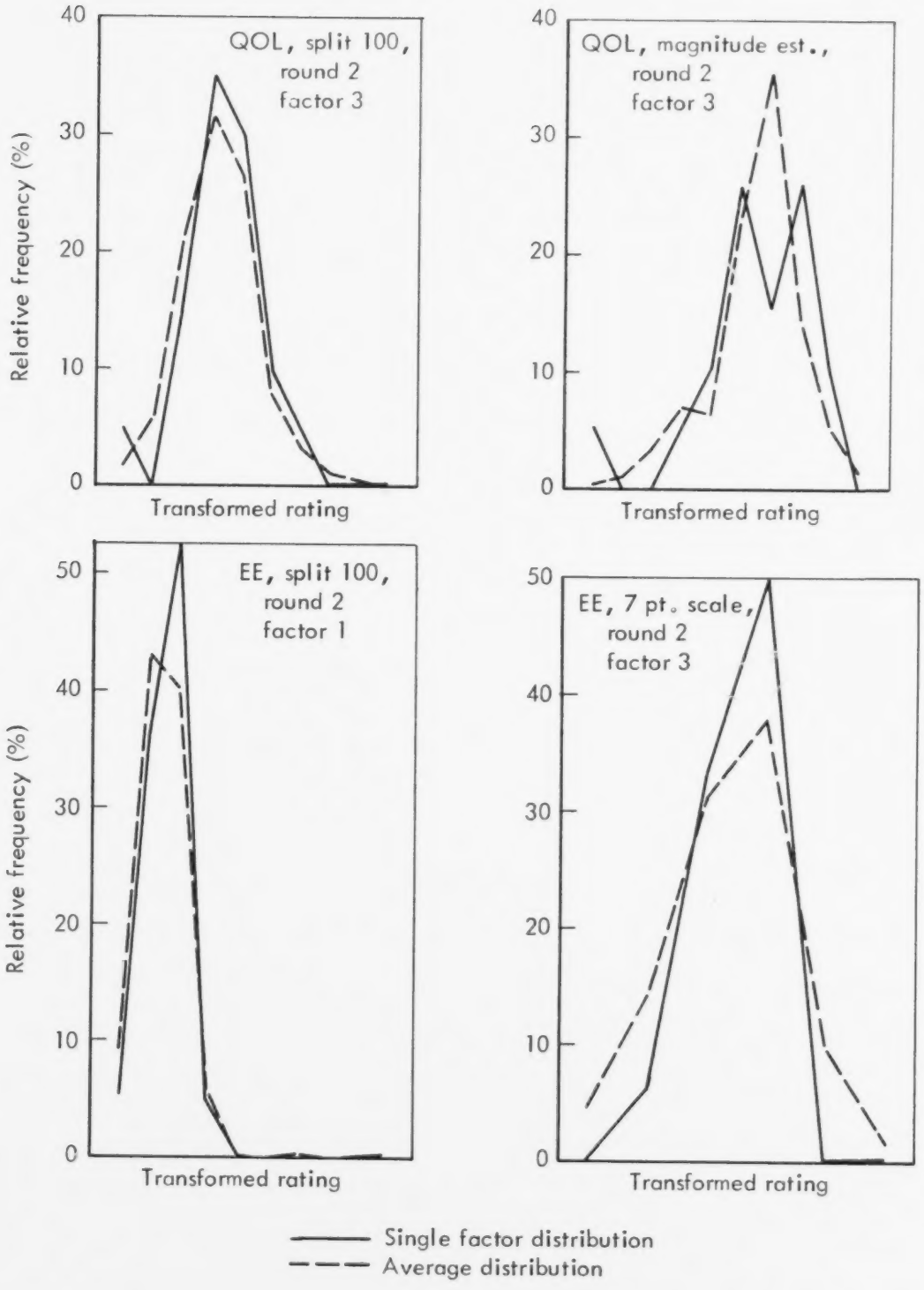

Fig.7—Representative distributions for single factors

point scale. Although the QOL group was more familiar with the QOL factors and the EE group with the EE factors, the relevance ratings from the two groups were in substantial agreement. The product-moment correlation between the two sets of ratings is 0.86. The mean ratings over the two groups combined are shown in Table 11.

The EE to QOL relevance ratings and the importance ratings of the QOL factors were used to determine the contribution of each of the EE factors to the quality of life in the following manner. Let e(i) be the contribution of the i'th EE factor to the quality of life, let r(i,j) be the relevance of the i'th EE factor to the j'th QOL factor, and let q(j) be the importance of the j'th QOL factor. The e(i) were computed as

$$e(i) = \sum_j r(i,j) \cdot q(j)$$

That is, the contribution of the i'th EE factor to the quality of life is the sum over all the QOL factors of the relevance of the i'th EE factor to the j'th QOL factor weighted by the importance of the j'th QOL factor.

A set of the "reweighted" EE factors was computed with the combined EE to QOL relevance ratings. The importance ratings of the QOL factors which were used in the computation were the round 2 medians from the QOL split-100 group. The reweighted EE factors are shown in Table 12; the entries in the table have been normalized to sum

Table 11

COMBINED MEAN RELEVANCE RATINGS, QOL AND EE GROUPS

EE Factors	Quality of Life Factors												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	2.671	1.566	1.237	1.118	1.131	2.250	1.803	0.671	1.171	1.421	1.776	2.447	1.473
2	2.592	1.790	1.513	1.171	1.013	2.303	2.210	0.579	1.947	1.895	1.526	2.211	1.631
3	1.618	1.671	2.158	0.973	1.289	1.895	2.026	0.539	2.671	2.395	1.236	1.750	1.185
4	0.947	1.395	1.315	2.065	1.855	2.000	1.763	0.579	1.105	0.645	2.724	1.145	0.315
5	1.224	1.447	1.750	1.053	1.724	1.539	1.447	0.869	2.474	1.487	1.500	1.487	0.737
6	1.145	2.434	1.895	1.369	1.368	1.527	2.855	1.197	1.605	2.448	1.250	2.276	2.053
7	0.776	1.408	1.013	0.671	1.381	1.000	1.039	0.658	1.394	0.934	1.118	0.948	0.408
8	1.500	2.368	1.947	1.645	1.566	1.895	2.750	1.105	1.777	2.461	1.750	2.250	2.500
9	1.737	1.816	1.750	1.013	1.237	1.447	2.039	0.474	2.171	2.289	1.158	1.948	1.408
10	1.816	1.803	2.605	1.382	1.671	1.724	1.987	0.803	2.342	2.552	1.645	1.684	1.803
11	1.079	1.342	1.671	1.039	1.184	1.618	1.776	0.619	2.566	2.026	1.263	1.421	0.881
12	1.276	1.684	1.473	1.763	1.829	2.565	2.066	0.671	1.473	1.171	2.461	1.526	0.921
13	0.644	1.461	1.289	1.040	1.750	0.843	1.184	0.895	0.724	1.053	1.013	1.092	0.974
14	2.013	1.632	1.316	1.224	1.158	2.289	2.131	0.763	1.474	1.632	1.921	1.829	1.039
15	0.750	1.131	0.685	0.711	1.092	1.250	0.855	0.421	0.723	0.500	1.079	0.724	0.250

Table 12

REWEIGHTED EE FACTORS

Factor (Listed as in Table 9)	Rank According to Reweighting	Reweighted Importanc Ratings (sum = 100)
Ability to learn	7	7.0
Broader outlook	4	7.8
Creativity	9	6.9
Social awareness	5	7.6
Communication skills	3	8.4
Tolerance	6	7.3
Self-awareness	2	8.5
Self-confidence	1	9.1
Responsibility	10	6.5
Impractical education	15	3.3
Career skills	12	5.7
Motivation	8	7.0
Involvement	11	6.2
Loss of idealism	14	4.2
Dependency	13	4.6

to 100 and are listed in the same order as in Table 9. The rank of each of the EE factors according to the reweighting is also given. The factor indices are keyed to the list of EE factors given in Table 4.

4. DISCUSSION

The results of applying the three criteria mentioned in the Introduction to the ratings of the educational and quality of life factors are all favorable to the hypothesis that Delphi procedures are appropriate for formulating group value judgments. The results with value material are in general comparable with factual material. This comment, however, must be taken with a certain amount of caution. The variability of performance on factual questions is large, depending on the type of question, and it is not entirely clear what would be an appropriate population of factual questions to compare with the value judgments.

With this caveat in mind, some gross comparisons can be made: the correlations between the median split-100 ratings and magnitude-estimation ratings on the QOL factors is 0.90 on the first round; the correlation between the median split-100 and 7-point ratings for the EE factors is 0.88 on the first round. These compare with an average correlation of 0.85 for similar groups making factual estimates of general information [3].

For the magnitude estimation and 7-point ratings of QOL and EE items respectively, convergence (variance reduction) occurred on all items in Round 2 (Tables 6 and 7). For split-100 ratings on the two sets of items, convergence occurred on all but 2 and 3 items respectively. For a set of 80 factual questions, convergence occurred in

97 percent of the cases. However, there is a difference in the amount of convergence. In a set of 8 exercises involving short-range (3 - 9 month) predictions of "newsworthy" events, the average reduction in standard deviation was about 40 percent; for the value items in the present study, standard deviations decreased about 19 percent for magnitude-estimation and 7-point ratings and about 10 percent for split-100 ratings. It seems probable that the constraint of adding to 100 for split-100 ratings decreased the convergence, but still the variance reduction on feedback was about twice as great for factual questions.

With regard to distribution shape, the major feature to note is that all of the distributions for all rating methods were single peaked. In addition, only four distributions out of 112 failed the goodness of fit (to the average distribution) test at the 10-percent level. This compares very favorably with similar tests for 80 factual questions where roughly one-fourth of the cases failed the test of fit to an average distribution (log normal) at the 10-percent level.

With respect to the number of changes of opinion between Round 1 and Round 2, the proportion of those who changed their estimate varied from 34 percent for the EE group making 7-point ratings to 49 percent for the QOL group making magnitude-estimation ratings. This compares with 65-percent changes for four control groups (receiving only

median and quartile feedback as in the present experiment) on factual questions [4]. The number of changes is lower for the value questions, but not so low as to reject the hypothesis that the subjects are responding to the feedback information.

Correlations were computed between the distance a subject's response was from the median on the first round, and the amount of change of the subject's response on the second round. These correlations are:

QOL, Split 100	0.40
QOL, Magnitude Estimation	0.41
EE, Split 100	0.54
EE, Magnitude Estimation	0.44

No comparable correlations have been computed for the data on factual questions; however, these correlations appear to be in line with the result [1] that for deviations from the mean of two quartiles, or less, the likelihood of a subject changing his estimate is roughly linear with deviation.

With the exception of the effects of iteration and feedback, the data generated by these experiments are similar to, and very much in line with, results obtained in a large number of experiments with psychophysical scaling, and with scaling "subjective magnitudes." The subjective magnitude scaling experiments, in fact, can be interpreted

as lending support to the general conclusion presented here. It is worth noting that the linear relationship between magnitude estimation and split-100 scaling indicated in Fig. 3 is in accord with the conclusion of S. S. Stevens [9] that ratio scales are relatively easy to obtain for a wide variety of subjective magnitudes with group estimation.

In the psychophysical and subjective magnitude studies, the role of the group judgment as opposed to individual judgments is left somewhat unclear. Stevens discusses the issue with respect to psychophysical judgments in terms of the similarity between individual intensity functions and group intensity functions. His assessment is that group judgments behave in the same general way as individual judgments. However, from the point of view of the present investigation, we are not so much concerned with the specific relationship of individual judgments to group, as we are to the assessment of the excellence of the group judgment. We take it for granted that individual judgments on both factual and value questions are based on incomplete, possibly biased, information; the general question, then, is to what extent pooling the judgments of a group of individuals is an improvement over the individual judgments. In the case of factual judgments of the sort studied in our experiments, the improvement is significant — overall group judgments were 45 percent more accurate than individual judgments. The present experiments (as well as the psychophysical ones) are compatible with the assumption that group judgments

are, on the whole, more "correct" for "subjective" judgments.

The effects of iteration and feedback — reduction in variance on the second round, and changes in scale values — are apparently new phenomena in the field of subjective magnitude scaling and psychophysical experiments. But they are not completely foreign to a related field of research, the study of attitude change. There do not appear to have been any experiments in attitude research concerning the results of feed-back of the simple sort we employed in the present experiments, but there is a large body of literature concerning what could be called feed-in of various kinds of material.* The focus of these experiments has been more on the phenomenon of change in attitudes and its determinants than on the question whether (in some sense) the changed attitudes were improved. However, one general consideration coming out of these studies is directly relevant: by utilizing various sorts of feed-in, much larger changes than we obtained with the statistical feed-back are easily obtained.

From the point of view of advancing the study of individual well-being or evaluation of higher education,

*We consider the experimental procedure employed by Asch [11] and others to be of this sort, although the information provided is generally misinformation; furthermore, the misinformation is presented so as to maximize the pressures towards conforming to the group response.

these exercises should probably be considered exploratory. The list of Quality-of-Life factors is similar to, but not identical with lists that have been generated in other exercises using different groups of respondents and somewhat different aggregation techniques [12]. The importance ratings are also similar to, but not identical with, importance ratings in the other exercises. To what extent these reflect differences in the manner in which the exercises were conducted, and to what extent differences in the life conditions of the respondent groups cannot be evaluated at present. Studies by Rokeach [5, 13] and others have shown that there are major differences in the ranking of terminal values depending upon income, education, and other characteristics of respondents. There is no inconsistency between assuming a fair amount of stability for basic value categories and varying importance ratings on these categories for different life states, if it is assumed that tradeoffs between basic values are meaningful, and depend on the state of the individual in the QOL space [12, Sec. II]. However, the present exercise was not sufficiently rich to test this hypothesis, nor do we know of any studies that have examined the question.

Nevertheless, several suggestive results have emerged from the present study. The most interesting is the large disparity in rank order of educational categories obtained from direct ratings and the rank order derived from the

weighted sum of judged contributions to the set of quality of life factors (Table 12). The very large shifts—cognitive skills moving from rank 1 to rank 7, creativity from rank 3 to rank 9, self-confidence from rank 8 to rank 1, etc.—are certainly formally significant. The result suggests as an interesting hypothesis for further exploration that some of the present discontent with the university stems in part from a (perhaps fuzzy) perception of just this disparity on the part of many students.

Another suggestive result is the high rating students give to security and peace of mind. A well-worn comment in news media is that one trouble with students is they take affluence and security for granted, and thus are not firmly guided by the reality principle. These results would suggest perhaps the opposite is the case. Security is high in their list of values. Of course, the student judgments may concern a different conception of "security" than that envisaged by the news media.

REFERENCES

1. Dalkey, N. C., The Delphi Method: An Experimental Study of Group Opinion, The Rand Corporation, RM-5888-PR, June 1969.

2. Brown, B., S. Cochran, and N. C. Dalkey, The Delphi Method, II: Structure of Experiments, The Rand Corporation, RM-5957-PR, June 1969.

3. Brown, B., S. Cochran, and N. C. Dalkey, The Delphi Method, III: Use of Self-ratings to Improve Group Estimates, The Rand Corporation, RM-6115-PR, November 1969.

4. Brown, B., S. Cochran, and N. C. Dalkey, The Delphi Method, IV: Effect of Percentile Feedback and Feed-in of Relevant Facts. The Rand Corporation, RM-6118-PR, March 1970.

5. Rokeach, M., Beliefs, Attitudes and Values, Jossey-Bass, Inc., San Francisco, 1968.

6. De Jouvenel, Bertrand, The Art of Conjecture, Basic Books, New York 1967.

7. Martino, Maj. J., "The Paradox of Forecasting," The Futurist, February 1969.

8. Johnson, S. C., "Hierarchical Clustering Schemes," Psychometrika, Vol. 32, No. 3, 1967, pp. 241-254.

9. Stevens, S. S., "Ratio Scales of Opinion," in D. K. Whitla (ed.), Handbook of Measurement and Assessment in Behavioral Science, Addison-Wesley, Reading, Massachusetts, 1968.

10. Blalock, Hubert M., Social Statistics, McGraw-Hill, New York, 1960.

11. Asch, S. E., "Effects of Group Pressure Upon the Modification and Distortion of Judgments," in E. E. Maccoby, T. M. Newcomb, and E. L. Hartley (eds.), Readings in Social Psychology, Henry Holt, New York, 1958, pp. 174-183.

12. Dalkey, N. C., R. Lewis, and D. Snyder, Measurement and Analysis of the Quality of Life, The Rand Corporation, RM-6228-DOT, August 1970.

13. Rokeach, M., and S. Parker, "Values as Social Indicators of Poverty and Race Relations in America," The Annals of the American Academy of Political and Social Science, Vol. 388, March 1970, pp. 97-111.